it is usually not necessary to invoke drastic *ad hoc* hypotheses and oversimplifications, which must be made in many other areas of physiological analysis. Finally, and most important, the physiological significance of the vascular system means that selection pressure must operate strongly on vascular design, and therefore according to Section 1.3, that optimality arguments are directly applicable to this system.

### 3.2 Vascular Branching

One of the characteristics of the vascular system, particularly in the arteries and capillaries, is the enormous amount of branching

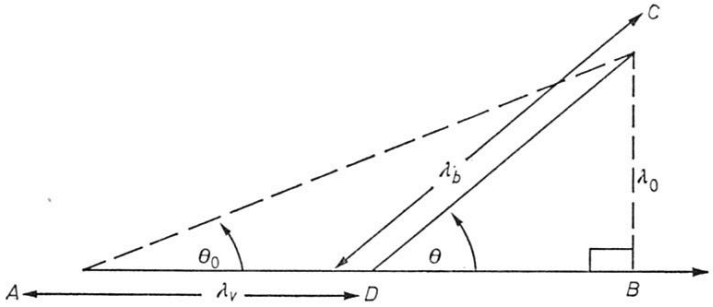


*Figure 3.1*

of vessels. If one attempts to follow the progress of a single vessel, one finds that it repeatedly gives off branches and bifurcations before it enters the venous system which conducts blood back to the heart. One of the most accessible applications of optimality to vascular structure has to do with questions concerning the angle between a vessel and a branch to which it gives rise, or with the relative diameters of main vessel and branch. In this section, the optimal angle at which a small vessel branches off a larger one, with respect to a plausible cost functional, is computed.

Suppose, as in *Figure 3.1*, that the main vessel of radius $r_0$, runs from the point $A$ to the point $B$. We further suppose that it is necessary to supply some point $C$, not on the line determined by



$A$, $B$, with a branch vessel. The radius of this branch vessel is assumed to be given; call it $r_1$.

The question is: at what point along the vessel $AB$ will the branch arise?

If one thinks for a moment about the flow of fluid through a system of conduits of constant radius, one recognizes that a certain amount of work must be done to push a given volume of fluid through the system. The amount of work required may be measured by the resistance of the system; the greater the resistance of the system to flow, the greater the amount of work which must be done in pumping fluid through it. Dealing with a vascular system, through which blood is pumped by a heart running on metabolic energy, it seems evident that the lower the resistance of the vascular system, the more efficient that system will be. With respect to the particular problem being considered, the Principle of Optimality suggests that the branch vessel should arise at that point $D$ which minimizes the resistance of the path $ADC$. In other words, the cost functional involved in this first simple approach to vascular system design is simply the total resistance of the system to flow.

Intuitively, it can be seen that the resistance of a system of conduits will depend directly on the total length of the system; the greater the distance the fluid has to flow, the greater is the resistance. It is also easy to see that the resistance decreases inversely with the radii of the conduits involved; the smaller the radii, the greater the resistance. Turning to *Figure 3.1*, it is seen that the total length of the system can be made as small as possible by originating our branch at $A$. However, this would have the effect of diminishing the overall average radius of the system. On the other hand, the system will have the largest possible average radius when the branch $DC$ arises perpendicularly to the main vessel, but this of course greatly increases the total length of the system.

To solve the problem with respect to the cost functional chosen, let $R_v$ represent the resistance per unit length of the main vessel $AB$, and let $R_b$ be the resistance per unit length of the branch. The total resistance $R_T$ of the path $ADP$ is then just

$$R_T = R_v\lambda_v + R_b\lambda_b \tag{3.1}$$

where $\lambda_v$ is the length of $AD$ and $\lambda_b$ is the length of $DC$. Using

elementary trigonometry, it is found that, in terms of the angle $\theta$,

$$R_T = R_v\lambda_0(\cot\theta_0 - \cot\theta) + R_b\lambda_0 \operatorname{cosec}\theta \tag{3.2}$$

Now according to Poiseuille's law[2], which governs the flow of fluid in rigid pipes, the resistances $R_v$, $R_b$ are related to the respective radii by the expressions

$$R_v = kr_0^{-4};\ R_b = kr_1^{-4}$$

where $k$ is a constant of proportionality (involving the viscosity and density of the fluid and the length of the pipe). Substituting these relations into (3.2), one obtains an expression for $R_T$ as a function of the angle $\theta$:

$$R_T(\theta) = \text{constant} + k\lambda_0\left(\frac{\operatorname{cosec}\theta}{r_0^4} - \frac{\cot\theta}{r_0^4}\right) \tag{3.3}$$

To find the value of $\theta$ which minimizes (3.3), one need only differentiate (3.3) and equate the result to zero, and find that the minimum angle $\theta_{\min}$ is given by

$$\theta_{\min} = \arccos\,(r_1^4/r_0^4) \tag{3.4}$$

*Exercise*

Verify that the angle $\theta_{\min}$ is a true minimum.

### 3.3 Same Problem: New Cost Functional

Before comparing the result (3.4) with empirical observations, proceed further in the analysis. It is assumed, in deriving (3.4), that it is only the total resistance to flow which has to be minimized. However, it is known that there is a certain amount of work which the organism must expend in the maintenance of any particular anatomical structure, and it suggests itself, in accord with the discussion in Section 1.4, that this work of maintenance should be included in the cost functional which is being minimized. Intuitively, it is reasonable to suppose that the work of maintenance associated with a portion of the vascular system would depend directly on the volume of the system.

Now the volume of the vessel $ADC$ is not minimized by the angle $\theta_{\min}$ given by (3.4) above.



*Exercise*

Find the angle which does minimize the volume of $ADC$.

Therefore, a cost functional which depends on the total resistance and the volume of the vessel $ADC$ will not be minimized by the expression (3.4). However, carrying through the computation for the sake of the present argument, the cost functional is now given by the expression

$$S_T = R_T + K(\lambda_v\pi r_0^2 + \lambda_b\pi r_1^2) \tag{3.5}$$

Here $K$ is a constant of proportionality and the bracketed expression is, of course, the volume of the path $ADC$.

Proceeding just as in the last section, it is found that the optimal branching angle $\theta_{\min}$, relative to the cost functional $S_T$, is given by

$$\theta_{\min} = \arccos\frac{r_1^4}{r_0^4}\cdot\frac{k + Kr_0^6}{k + Kr_1^6} \tag{3.6}$$

Note that in the limiting case $r_1 \to r_0$, the expression (3.6) becomes identical with (3.4). Likewise, the limiting situation $r_1 \to 0$ gives identical results with those obtained from (3.4). Between these two limiting cases, however, the formulas (3.4) and (3.6) do not agree.

How well do the simple arguments above compare with what is observed in a real vascular system? In particular, is (3.1) or (3.5) the more realistic cost functional? The literature contains a number of empirical laws on branching in the vasculature, ascribed to Wilhelm Roux, and which are cited by d'Arcy Thompson[3] as follows:

(1) If an artery bifurcates into two equal branches, these branches come off at equal angles to the main stem.

(2) If one of the two branches be smaller than the other, then the main branch, or continuation of the original artery, makes with the latter a smaller angle than the main branch.

(3) All branches which are so small that they do not diminish the main stem come off from it at a large angle.

(The reader may find it interesting to compare Hess' argument establishing the result (3.4) cited in d'Arcy Thompson, *loc. cit.*, with the one presented above.)

It should be noted that the only one of these rules which is strictly in accord with the hypothesis made is the third, and this is

obviously compatible with (3.4) and with (3.5). However, the hypothesis is restricted entirely to *branching* of vessels, while the first two of these rules refer to *bifurcation*; *Figure 3.2* will make clear the difference between these two notions.

Thus, the derivations both accord with observation as far as they go, but they must be enlarged to consider the case of bifurcation before some understanding of the way vessels subdivide in the vascular system can be assumed. This means essentially that the effect of the segment *DB* in *Figure 3.1* must be taken into account.

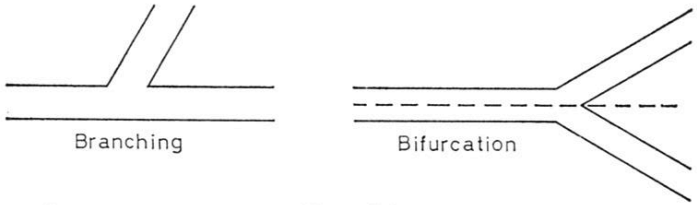


*Figure 3.2*

### 3.4 A Note on Units and Dimensions

It may be well to pause for a moment to consider the physical significance of the cost functionals that have been introduced. In order to do this, the concept of the *physical dimension* of the various physical quantities involved should be explained to those unfamiliar with this notion.

All measurements in mechanics ultimately reduce to measurements of length, mass, and time, and consequently every physical quantity arising in mechanics can be expressed in a unique way from a sequence of measurements of mass, length, and time alone. For instance, velocity is defined to be length per (unit) time; acceleration is velocity per (unit) time, or length per time per time; force is mass times acceleration or mass times length per time per time, etc.

The basic entities, mass, length, time, are called the fundamental mechanical dimensions, and the dimension of every mechanical quantity is some monomial of the form $m^a l^b t^c$, where



$a, b, c$ are rational integers. Note that the dimension of a physical quantity remains the same regardless of the particular set of units used to measure the fundamental dimensions: velocity is length per time whether measured as centimeters per second or miles per hour.

In an equation between physical quantities it is impossible to add quantities of unlike dimension. Thus, a velocity can be added to a velocity, and a work to a work; but a velocity cannot be added to a work. Moreover, a physical equation must be dimensionally homogeneous; the dimensions of the expressions on both sides of the equation must match. Dimensional analysis, which treats these matters systematically, is a rather interesting study in itself, but it will not be pursued further here[4].

Now let us look at our cost functionals (3.1) and (3.5). The cost function (3.1) is a resistance to flow, and the dimension of resistance (in terms of $m$, $l$, $t$) is $ml^{-4}t^{-1}$. However, the functional (3.5) is not so obvious. The first term is again a resistance, but the second term is proportional to a volume; i.e. to a quantity whose dimension is $l^3$. However, a dimension may be assigned to the constant of proportionality *ab lib.*; in order to insure dimensional homogeneity, it is necessary that the constant has the dimension $ml^{-7}t$. If this is carried out, the cost functional (3.5) also assumes the character of a resistance.

However, it will be convenient in what follows to alter the significance of the cost functional (3.5), and to consider, instead of the resistance of the system, the power dissipated during flow, i.e. the work done per unit time by the flow. Now power has dimension $ml^2t^{-3}$. Resistance can be converted into power by multiplying by a factor with dimension $l^6t^{-2}$. It should be observed that this last dimension is just the square of the dimension of flow; flow $= l^3t^{-1}$. Thus it is always possible to obtain the power dissipated in a hydraulic system from its resistance by multiplying by the square of the flow; in particular, the first term of (3.5) can be converted to the dimension of power. The second term may be so converted, without change of form, by simply altering the dimensional significance of the constant $K$. In the present case, it is readily verified that $K$ must have the dimension $ml^{-1}t^{-3}$. If this is done, then (3.5) assumes the form

$$P_T = f^2 R_T + K(\lambda_v \pi r_0^2 + \lambda_b \pi r_1^2) \tag{3.7}$$

and it may be noted that, in any particular instance in which the flow $f$ is given, the unit of flow may be chosen so that $f^2=1$. In that case (3.7) becomes formally identical with (3.5), although the physical and dimensional significance of the two expressions are, of course, quite different. However, the reader will verify that even in the apparently more general form (3.7), the result remains formally unchanged; the constant $k$ need only be replaced by $f^2k$.

### 3.5 Vascular Bifurcation

Returning to the main argument, the cost functional will be taken to be of the form (3.7) which has just been derived. This allows the use of an elegant argument, due essentially to Murray[5], and the by-passing of the rather formidable calculations involved in a direct attack on the bifurcation problem.

The first thing to note is that, from (3.7), the optimal radius of a vessel of fixed length can be calculated for a given flow. For a single unbranched vessel of length $L$, the cost functional (3.7) becomes

$$P_T = f^2 R_T + K\pi r^2 L \tag{3.8}$$

This is differentiated with respect to $r$ and the result equated to zero; it is then found that $P_T$ attains its minimal value when

$$f^2 = \frac{r^6 \pi K L}{2k} \tag{3.9}$$

Substituting (3.9) in (3.8) at the optimal flow rate $f$, the power dissipation per unit length is related to the radius of the vessel by the expression

$$P_T/L = \gamma r^2 \tag{3.10}$$

where $\gamma$ is a constant.

Next, suppose that the bifurcation which minimizes (3.7) is known, and let it be that configuration determined by $A$, $B$, $C$, $D$ in the following figures. Then, from *Theorem 2.7*, it is known that a first-order change in the minimal configuration produces only a second-order change in the cost functional; that is, up to infinitesimals of second order, the value of the functional $P_T$ is unchanged if the minimal configuration is replaced with a 'neighbouring' configuration, differing from the minimal by a



first-order infinitesimal. Now consider the three following special 'neighbouring' configurations, displayed in *Figures 3.3*, *3.4*, and *3.5*. Here the line interval $\overline{BB'}$ is considered to be a first-order

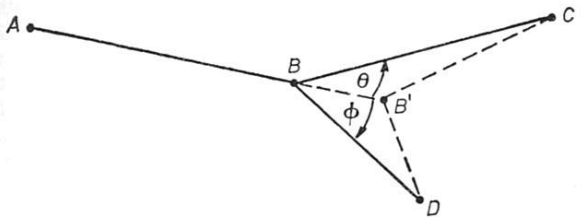


*Figure 3.3. From* C. D. MURRAY[16]

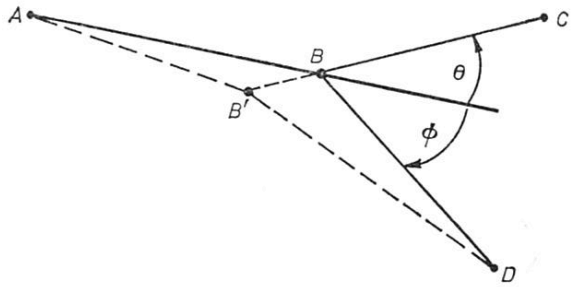


*Figure 3.4. From* C. D. MURRAY[16]

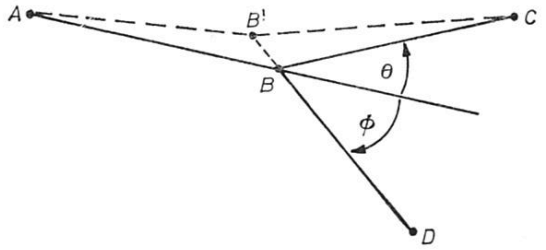


*Figure 3.5. From* C. D. MURRAY[16]

infinitesimal, which is denoted by $\delta$. The radii and lengths of the three limbs $AB$, $BC$, $BD$ of the minimum configuration are denoted by $r_0, L_0$; $r_1, L_1$; $r_2, L_2$ respectively.

Now consider *Figure 3.3*. The equilibrium path $AB$ has been perturbed by an increase in length of an amount $\delta$; on the other hand, up to second-order quantities, the equilibrium paths $BC$ and $BD$ have been correspondingly shortened by the amounts $\delta\cos\theta$, $\delta\cos\phi$, respectively. If there is an optimal flow [i.e. the flow and the radius are related by the expression (3.9)], then the increment in the 'cost' accruing to the perturbed paths $AB'$, $B'C$, $B'D$ are just $-r_0^2\delta$, $r_1^2\delta\cos\theta$, $r_2^2\delta\cos\phi$, respectively. But as pointed out, the sum of these increments (up to the second order of small quantities) must vanish; i.e.,

$$r_0^2\delta = r_1^2\delta\cos\theta + r_2^2\delta\cos\phi \tag{3.11}$$

Performing the same computation on the configurations shown in *Figures 3.4* and *3.5*, one obtains respectively the relations

$$r_1^2\delta = r^2\delta\cos\theta - r_2^2\delta\cos(\theta+\phi) \tag{3.12}$$

$$r_2^2\delta = r_0^2\delta\cos\phi - r_1^2\delta\cos(\theta+\phi) \tag{3.13}$$

Solving the three equations (3.11), (3.12), and (3.13) for the three angles $\theta$, $\phi$, $(\theta+\phi)$, one finds the relations

$$\cos\theta = \frac{r_0^4 + r_1^4 - r_2^4}{2r_0^2 r_1^2} \tag{3.14}$$

$$\cos\phi = \frac{r_0^4 + r_2 - r_1^4}{2r_0^2 r_2^2} \tag{3.15}$$

$$\cos(\theta+\phi) = \frac{r_0^4 - r_1^4 - r_2^4}{2r_2^2 r_2^2} \tag{3.16}$$

These equations may be simplified by invoking relation (3.9). It is noted that, at optimality, the flow $f$ is proportional to the cube of the radius of the vessel. Let the flow in the branches $AB$, $BC$, $CD$ be denoted by $f_0, f_1, f_2$, respectively. Evidently $f_0 = f_1 + f_2$; from this and (3.9) it follows that



$$r_0^3 = r_1^3 + r_2^3 \tag{3.17}$$

which is an important relation. Substituting (3.17) into (3.14), (3.15), (3.16), there is obtained

$$\cos\theta = \frac{r_0^4 + r_1^4 - (r_0^3 - r_1^3)^{4/3}}{2r_0^2 r_1^2} \tag{3.18}$$

$$\cos\phi = \frac{r_0^4 + r_2^4 - (r_0^3 - r_2^3)^{4/3}}{2r_0^2 r_2^2} \tag{3.19}$$

$$\cos(\theta+\phi) = \frac{(r_1^3 + r_2^3)^{4/3} - r_1^4 - r_2^4}{2r_1^2 r_2^2} \tag{3.20}$$

The reader should satisfy himself that these expressions do account for the qualitative laws of Roux which are mentioned above. A comparison of optimal bifurcation angles computed as above with actual angles observed in corrosion preparations of cat lung (cf. Murray, *loc. cit.*) shows a close agreement. Therefore, it may be concluded that the cost functional (3.7) represents a good approximation, at least, to the actual cost functional in real situations, and that arterial branching is indeed quantitatively understandable in terms of optimization with respect to that cost functional[6].

*Exercises*

The limiting situation characterized by $\phi = 0$, $r_0 = r_2$ corresponds to the branching of vessels considered in Section 3.3. Do the formulas (3.14) to (3.16) reduce to the result (3.6) obtained therein for the optimal branching angle? If not, can you explain the discrepancy?

Compute the optimal bifurcation angle using the cost functional (3.1). Compare your results with those obtained above, and in the limit $\phi = 0$, $r_0 = r_2$, with (3.4).

## 3.6 The Optimal Radius of a Branch

The problem of optimal radius is really solved, relative to the cost functional (3.7), by expression (3.17). There is, however, one

special case of interest, which deserves explicit mention. Suppose that the original vessel bifurcates into two vessels of equal radius; i.e. $r_1 = r_2$. It then follows from (3.17) that $r_0 = \sqrt[3]{2}\, r_1$, or

$$r_0 = \cdot 794\, r_1 \tag{3.21}$$

This is the result obtained by Cohn[7], using what seems at first sight to be a quite different argument. However, the interested reader may verify that the cost functional involved in Cohn's computation is equivalent to the one used above.

### 3.7 The Aortic Radius[7]

It is seen that, given the radius $r_0$ of a vessel, the radii of the branches and bifurcations which arise from it can be computed, if optimality is assumed with respect to a cost functional of the form (3.1) or of the form (3.7). Thus, given the radius of the initial vessel of the vascular system, i.e. of the aorta, one can in principle compute the radii of all the other vessels of the system. The question then arises: can the aortic radius on the basis of optimality considerations be directly computed?

First, one has to decide on the appropriate cost functional with respect to which the aortic radius is to be optimized, and one's immediate reaction is to use the same functional (3.7) which has been so successful in the computations carried out on vascular branching. However, it can readily be seen that this functional will not work in the computation of the aortic radius. It has already been noted from (3.9) that at optimality with respect to the cost functional (3.8), the radius of the vessel is proportional to the cube root of the flow. The constant of proportionality is given by

$$\frac{\sigma\pi^2}{16\eta} \tag{3.22}$$

where $\eta$ is the viscosity of the blood, $\sigma$ the density. The constant $\sigma/\eta$, dimensionally, is of the form $ml^{-1}t^{-3}$, which can be expressed also as $(ml^2t^{-2})\,(l^{-3}t^{-1})$; i.e., as energy per volume time; in the usual c.g.s. units, this constant has been evaluated by Murray[8] to be approximately 19,000 ergs/cm$^3$ sec. Using this value, and taking the viscosity $\eta$ to be ·03 *P* (the poise being the unit of viscosity; 1 *P* being equal to 1 dyne-sec/cm$^2$), the constant (3.22) is given



approximately by $3{\cdot}8 \times 10^6$ c.g.s. units. In man, the flow $f$ is given approximately by 100 cm$^3$/sec, so that according to (3.9)

$$(100)^2 = (3{\cdot}8 \times 10^6) r^6$$

from which $r \sim 0{\cdot}4$ cm. The correct value of the radius of the human aorta is approximately 1·5 cm, so one can see immediately that optimality of the aortic radius is not determined by the cost functional (3.8).

Qualitatively, it is not hard to see why (3.8) will be inapplicable to the aortic radius. The aorta receives the full output of the heart, and the velocity of blood flow in the aorta as blood is expelled from the heart is far higher than at any other point in the vascular system. At high velocities of flow, a condition arises which, in hydro-dynamics, is known as turbulence, in which the simple flow-pressure relations that have been used no longer hold. In particular, a much greater amount of work must be done to move a quantity of fluid in turbulent flow than to move the same fluid in simple laminar flow, since turbulence dissipates a great deal of energy.

What is the condition for turbulent flow? With any flow of average velocity $v$ in a rigid pipe of radius $r$ there may be an associated dimensionless number called the Reynolds number:

$$\rho = \frac{\sigma r v}{\eta} \tag{3.23}$$

For each fluid, there is a critical value of the Reynolds number, above which turbulent flow sets in, and below which the flow remains laminar. This critical value cannot itself be computed from first principles, but must be determined experimentally; for blood, turbulence may ensue if the Reynolds number of the flow exceeds 1000 or thereabout.

The velocity of flow is inversely proportional to the cross-section area of the vessel. For a cylindrical vessel like the aorta, the velocity is thus inversely proportional to the square of the radius, and the Reynolds number (3.23) is inversely proportional to the radius itself. Thus, a small aortic radius is incompatible with a low Reynolds number, i.e. with laminar flow in the vessel. In other words, simple minimization of resistance and maintenance are not sufficient to account for the aortic radius; the value

obtained above for the human aortic radius in terms of such a cost functional would result in an aortic flow which would be essentially turbulent. Qualitatively, this might be remedied in two ways: by building a larger heart, or by increasing the aortic radius. It is obviously far less drastic to increase the aortic radius above the value required by strict minimization of the cost of maintenance.

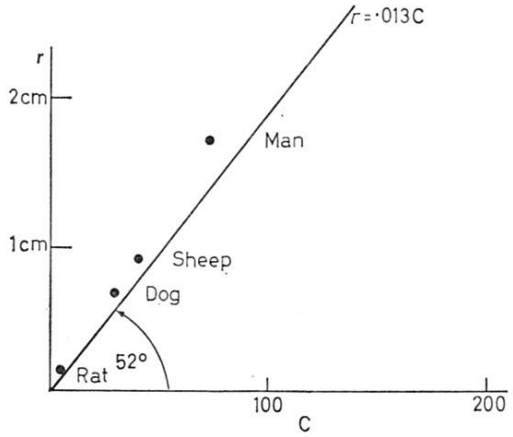


*Figure 3.6*

It is a simple matter to compute from the above the smallest value of the aortic radius compatible with laminar flow for a given cardiac output[9]. (This, it should be noted, is not a true minimum problem, but rather a 'threshold' one.) The average velocity $v$ of the flow in the aorta is defined to be

$$v = C/\pi r^2$$

where $C$ is the cardiac output and $r$ is the aortic radius.

Substituting this expression into (3.23), putting $\sigma/\eta = 37$ and $\rho \geqslant 1000$, a universal relation between aortic radius and cardiac output is found:

$$r \geqslant \cdot 013C \tag{3.24}$$



For man the cardiac output $C$ is about 100 cm³/sec; therefore from (3.24)

$$r \geqslant 1{\cdot}3 \text{ cm}$$

Since the human aortic radius is of the order of 1·5 cm, this is an excellent agreement considering the rough values that have been employed. Note also the close agreement between the graph of (3.24) and the points corresponding to the actual values of cardiac output and aortic radius in a variety of different organisms, displayed in *Figure 3.6*.

## 3.8 The Capillaries

A number of further indices of vascular structure may immediately be derived from the foregoing, and tested against empirical observation. Let us suppose, as a first approximation, that every vessel in the vascular system (except the aorta, of course) arises from a bifurcation of a larger vessel. Let us further assume that these bifurcations are optimal; i.e. that the radii of the larger and smaller vessels involved in a bifurcation satisfy (3.21). Knowing that, in mammals, the radius of a capillary is about $5 \times 10^{-4}$ cm, it is therefore pertinent to ask for the number $n$ of bifurcations necessary to pass from an aorta of radius $r_a$ to a capillary of radius $5 \times 10^{-4}$ cm by successive bifurcations.

From assumptions, it readily follows that the desired number $n$ satisfies the relation

$$(\cdot 794)^n r_a = 5 \times 10^{-4} \tag{3.25}$$

For the dog, one may take $r_a = 0{\cdot}5$ cm. Substituting in (3.25) and solving for $n$, it is found that $n$ is about 30.

The assumption that each vessel in the vascular system arises by the bifurcation of a larger vessel implies that the total number of vessels in the vascular system is just

$$2^{30} \cong 10^9$$

The total number of capillaries, which according to the model are the vessels arising from the final bifurcation, is just half this number, or $0{\cdot}5 \times 10^9$. Data cited by Rashevsky[10] puts the number of capillaries in the dog, as obtained from empirical measurement, at $1{\cdot}2 \times 10^9$, which is already an excellent agreement. The agreement becomes even better when it is realized that

it will be difficult to distinguish empirically between the true capillaries (i.e. the vessels arising from the final bifurcation) from vessels arising from, say, the twenty-ninth or twenty-eighth bifurcation.

In the preceding discussion, it was supposed that the actual capillary radius was given on the basis of experimental observation. However, this radius too can be directly derived from optimality considerations. The fundamental idea behind such a derivation is that the capillary radius is limited by the rates of diffusion of essential metabolites into the surrounding tissues. The actual arguments involve no new principles, and since they are somewhat lengthy, the reader is referred to the original literature for full details[10].

### 3.9 A Note on the Erythrocyte[11]

The mammalian erythrocyte, or red blood corpuscle, is in itself a little textbook on biological optimality. This text can be read rather easily, at least as regards its main features, and it seems appropriate to include in the present chapter a discussion on optimality in the vascular system. The treatment given below is a qualitative one; a full mathematical discussion of the kind seen in the preceding sections would not be difficult, but would be quite lengthy.

First, there are some generalities about the form and structure of mammalian erythrocytes. The human erythrocyte, which may be regarded as typical, is basically a circular disc, with a diameter of approximately $8 \times 10^{-4}$ cm, and a thickness of $2 \times 10^{-4}$ cm. The two faces of the disc are not planar, but are closer to each other in the centre than at the edges of the erythrocyte, giving the cell its characteristic biconcave form. Erythrocytes are present in enormous numbers in the blood stream, there normally being between 4,000,000 and 6,000,000 cells in each cubic millimeter of blood.

The erythrocyte is, of course, primarily involved in oxygen transport and exchange of gases between the blood and the surrounding tissues. The basic instrument of oxygen transport and exchange is the protein haemoglobin. Each erythrocyte contains 30–35 per cent by weight of haemoglobin; 60–65 per cent of the cell is water, and the small remaining portion of the cell is comprised of lipoprotein, a few enzymes, and small molecules



arranged in a structure which has not yet been completely clarified. Unlike all other cells, including the erythrocytes of non-mammalian vertebrates, mammalian erythrocytes lack a nucleus.

To best appreciate the various optimality features of the erythrocyte, imagine that one is engaged on the design of a circulatory fluid which will be of optimal efficiency in gas transport and exchange. According to the Principle of Optimal Design, this ideal circulatory fluid should be very close to what is actually found in mammalian blood.

The very first question in designing such a circulatory fluid is: Should the ideal fluid contain blood cells at all? *A priori*, it seems inefficient to enclose all the haemoglobin within cells, thereby interposing one more barrier (the cell membrane) to the free diffusion of gases to and from the tissues. Would not a blood stream in which haemoglobin appears free in solution in the plasma—as in many invertebrates—be more efficient in oxygen transport and exchange? A simple computation shows that this cannot be the case. For one thing, if the amount of haemoglobin required to meet the metabolic requirements of a human being were dissolved in a volume of fluid equal to the normal plasma volume, the resulting solution would be exceedingly viscous. To keep such a viscous solution in circulation would require a very large expenditure of energy, i.e. a large heart. Further, whenever a viscous fluid flows in cylindrical vessels, the friction between the walls of the vessel and the fluid layer adjacent to the walls tends to immobilize that layer. Thus, the circulatory motion would be confined to the fluid in the centre of the blood vessels; the plasma adjacent to the vessel walls would tend to stagnate, thereby nullifying whatever advantage in gas exchange might be gained by freeing the haemoglobin in the plasma. Accordingly we conclude that free haemoglobin in the plasma of actively metabolizing organisms is grossly inefficient, and that the haemoglobin must be packed into specific formed elements. Hence, optimality considerations imply the very existence of erythrocytes in such organisms.

We must now determine the optimal shape of the formed elements whose existence has already been determined. Briefly, the following factors must be taken into consideration:

(1) For optimal efficiency in gas exchange, the formed elements must possess a maximal surface area per unit volume.

(2) The formed elements must be such as to flow with a minimal dissipation of energy; e.g., they must make a minimal disturbance of the lines of flow of plasma through the vascular system.

(3) The formed elements must not settle out and thereby occlude the blood vessels; i.e. there must be a maximal drag on them by the circulating plasma.

(4) The formed elements must be comparable in size with the smallest vessels in the vascular system.

The statement (1), that an optimal formed element must possess a maximal surface area per unit volume, bears a superficial resemblance to the isoperimetric problem[12], one of the classical problems of the calculus of variations. However, it can generally be seen that the stated problem is one for which no solution exists (cf. Section 2.7); a solid may be deformed so as to have an indefinitely large surface area while maintaining a constant volume. On the other hand, it is apparent, and it can readily be proved that a discoidal solid will possess a considerably greater surface area than a sphere of equal volume. From direct measurements on actual erythrocytes, it has been estimated that the gain in surface area is of the order of 30 per cent. Hence it is clear that it is necessary for the formed elements to be discoidal rather than spherical in shape (even though energy seems to be required to maintain the discoidal configuration).

Regarding the biconcavity of the typical mammalian erythrocyte, Hartridge[13] pointed out many years ago that '. . . the most efficient shape for the corpuscle would seem to be that which allows the whole of the contents to be saturated in the same interval of time, and not that in which the edge would be saturated before the centre and vice versa.' Now the edges of a disc with planar faces are exposed to the environment on three sides, while the centre is only exposed on two, so that in such a disc gaseous diffusion involving molecules at the centre of the disc volume is more difficult than for molecules near the edges. The only way to avoid this state of affairs is to have the centre of the disc thinner than at the edges; the resulting form is a very close approximation to the biconcavity of the actual mammalian erythrocyte.

It is not hard to show rigorously, on hydrodynamic considerations, that the requirements (2), (3) are already implied by the discoidal form which satisfies (1). For instance, it is known[14] that



a suspension will flow through a tube with minimal energy dissipation if the suspended particles are concentrated around the axis of the tube (i.e. in the region of minimal shear). It can be shown that a suspension of discoidal particles will exhibit this axial accumulation, thereby minimizing both the energy of flow and the tendency for settling out; and furthermore, a suspension of discoidal particles will exhibit these properties to a greater degree than will a suspension of spheres of equal volume.

Finally, from (3.25) there is an upper bound on the volume of the hypothetical ideal formed element. Since ideally each corpuscle should carry as much haemoglobin as possible, it is intuitively clear that the size of the corpuscle should be as close as possible to the maximal allowable: i.e. should possess a radius of approximately $10^{-4}$ cm[15].

The similarity between the ideal circulatory fluid and the actual mammalian circulatory fluid is thus exceedingly close. The correspondence of prediction with observation, which has been observed throughout the present chapter, shows clearly how useful the Principle of Optimal Design can be in accounting for, and in predicting, important aspects of anatomical form.

## Notes to Chapter 3

1. A corrosion preparation of an organ (e.g., lung, liver, kidney) is obtained by injecting the system through a major artery with a coloured liquid plastic under pressure. The plastic is allowed to harden, and the soft tissue dissolved away with dilute acid. The result is an exact cast of the vascular tree of the organ.

2. Poiseuille's law was discovered empirically, but can be derived from elementary hydrodynamic principles. See for instance G. Joos, *Theoretical Physics* (2nd edn), Hafner, New York, 1950, pp. 212–13.

3. Cf. D'Arcy W. Thompson, *On Growth and Form* (revised edition), Macmillan, New York, 1945, pp. 948–57, especially pp. 951–3.

4. See, for example, P. W. Bridgman, '*Dimensional Analysis*' (revised edition), Yale University Press, 1956. For an attempt to apply dimensional methods directly to biological problems, the reader might consult W. R. Stahl, 'Dimensional Analysis in Mathematical Biology I', *Bull. math. Biophys.*, 23 (1961) 355–76; II, *ibid.*, 24 (1962) 81–108.

5. C. D. Murray, 'The Physiological Principle of Minimum Work I', *Proc. nat. Acad. Sci.*, 12 (1926), 207–14; 'The Physiological Principle of Minimum Work Applied to the Angle of Branching of Arteries', *J. gen. Physiol.*, 9 (1926) 835–41.

6. Of course, in the case of the large arteries coming off the aorta and supplying major organs, other anatomical considerations take precedence over the minimization of (3.7). However, with regard to secondary

arteries and arterioles, the vast majority of which are subject to no anatomical exigencies, the stated result holds.

7. D. Cohn, 'Optimal Systems: I. The Vascular System', *Bull. math. Biophy*, 16 (1954) 59–74; 'Optimal Systems II: The Vascular System', *ibid.*, 17 (1955) 219–27.

8. C. D. Murray, 'The Physiological Principle of Minimum Work II', *Proc. nat. Acad. Sci.*, 12 (1926) 299–304.

9. Actually, some turbulence seems to inevitably occur in the aorta during particular portions of the cardiac cycle. See, for example, J. MacDonald, 'The Occurrence of Turbulent Flow in the Rabbit Aorta', *J. Physiol.*, 118 (1952) 340–7. Such observations, however, do not greatly affect our arguments concerning aortic design.

10. Cf. D. Cohn, *op. cit.*, N. Rashevsky, *Mathematical Biophysics*, Dover, New York, 1960, Vol. II, Chap. XXVII.

11. An excellent review article on the comparative biology of the erythrocyte, with a fine bibliography, is the article by H. Lehmann and R. G. Huntsman, 'Why are Red Cells the Shape They Are?', which comprises Chap. 2 of the book *Functions of the Blood*, R. G. MacFarlane and A. H. T. Robb (eds.), Academic Press, New York, 1961. Chaps. 1, 4, 8, 9 are also germane to our discussion.

12. The simplest form of the isoperimetric problem is to determine that plane curve of given perimeter which encloses a maximal area. The solution, of course, is the circle.

13. Hartridge, H. 'The Shape of the Red Blood Corpuscle', *J. Physiol.* 53 (1919) 81–85.

14. See, for example, A. Jefferey, *Proc. roy. Soc.* A, 107 (1922) 161. Also the article by L. E. Bayliss, 'Rheology of Blood and Lymph' which comprises Chap. 7 of 'Deformation and Flow in Biological Systems', A. Frey-Wissling (ed.), Interscience, New York, 1952.

15. It was stated in Section 3.8 (cf. Note 10 above) that the capillary radius could itself be determined directly from optimality considerations. These considerations do not involve the size of the erythrocyte, so that the determination of erythrocyte radius on the basis of capillary dimensions does not involve a circularity.

16. Figures 3.3, 3.4 and 3.5 are reprinted by permission of The Rockefeller University Press from The Journal of General Physiology, 1926. Vol. 9, p. 836.

# 4

# ONTOGENY AND PHYLOGENY

## 4.1 Introduction

IN THE preceding chapter it has been shown how the Principle of Optimal Design might be applied to problems of biological structure. Such studies are obviously of great importance, but from the standpoint of the biologist they leave something to be desired. For the biologist is essentially concerned with the interactions which take place between individual biological structures, considered as parts of a functioning whole. In the previous work, the structures in which interest was shown are considered to be isolated from the organism of which they are a part.

However, the integration of diverse functions and structures into a smoothly functioning organism is also a province of the theory of optimality. The purpose of the present chapter is to introduce the reader to a number of important biological ideas with which he may not be familiar, and which underlie most of the developments of the next few chapters. This discussion is rather more discursive and qualitative than the preceding one, but the mathematical cudgels are again taken up in Chapter 5.

## 4.2 Some Basic Concepts: Ontogeny and Phylogeny

The biologist is deeply concerned with the manner in which organisms change in time. His particular concern is with those sequences of inexorable structural changes which seem to be determined by the nature of the organism, and which are commonly called development.

The most obvious kind of developmental processes are those which take place during the lifetime of a single biological individual. Such an individual begins life as a single cell, which is usually comparatively unstructured (except perhaps for a large amount of food material or yolk). Through a process of cell division, and a differential specialization of the cells of the resulting aggregate, the developing organism passes through a sequence

of stages which are characterized by abrupt and extremely complex changes of form and function. The final adult stage of this process is characterized, in higher organisms, by the ability to reproduce sexually; no further important form changes occur until the individual dies.

The kaleidoscopic process by which an adult organism develops from a single cell, often by rather devious paths, is called *ontogeny*, or *ontogenesis*. In many organisms, ontogeny falls into a number of well-defined stages. In insects and frogs, for example, there are long, relatively static larval stages (the caterpillar and the tadpole, respectively) in which the developing organism lives freely in the world, and except for the absence of reproductive ability[1], seems effectively to be an adult. Nevertheless, after a characteristic interval has elapsed, the larval organism undergoes a complete metamorphosis from which the true adult form, often completely different from the larval, emerges. The ontogeny of many other organisms, however, does not contain a larval stage at all.

The main thing to notice about ontogeny, at present, is that its course is complete over a relatively short time interval. The transition from fertilized egg to reproductive adult seldom, except perhaps for some trees, takes more than 20 years, and in most forms is completed in a matter of weeks or months.

Now when a biologist characterizes a species of organism, he invariably uses the features of the adult form of the organism as his guide. Thus, one speaks of species of butterflies and frogs, not of caterpillars and tadpoles. The adult form of a species of organism stands at the end of a sequence of ontogenetic transitions, and this sequence, in a sense, constitutes the development of that form. But the theory of evolution has shown that there are other questions regarding the origin of the form of species, and so consideration is given to quite a different kind of development of the same organism.

The theory of evolution teaches, roughly, that each species has arisen as a result of morphological alterations in previously existing species and the operation of natural selection. Thus, it follows that any particular species can be traced backwards, through geological time, to the beginnings of life on the planet. An individual belonging to a particular species therefore stands also at the end of a second sequence of developmental transitions,



namely, the transitions of form which have occurred in the evolution of the species itself. This sequence of transitions is called *phylogeny*, or *phylogenesis*.

It is, of course, obvious that the ontogenetic and phylogenetic sequences involved in the development of a particular organism are in many ways very different from one another. Ontogeny, for example, can be studied directly; it can be watched in the laboratory or in a field. Phylogeny, on the other hand, is not usually amenable to laboratory treatment. Only rarely, and in very special situations, is it possible to directly observe phenomena resembling phylogenetic transitions[2]. Phylogeny is the province of the evolutionist and the Palaeontologist; ontogeny of the embryologist and, more recently, the biochemist.

Despite these differences, there are many striking similarities between the two developmental sequences. These are outlined in the next few sections, and then their relations discussed from the standpoint of optimality. These matters are dealt with in fuller detail in Chapters 5 and 6.

### 4.3 Recapitulation. The Biogenetic 'Law'

As soon as it was recognized that each organism stands at the end of two different developmental processes, it became natural to try to find some relation between the two. Such researches were spurred on by the fact that even the most cursory inspection of the ontogenetic and phylogenetic sequences reveals striking parallels between them.

First, look at phylogenetic sequences. Evolutionary and palaeontological theory indicate that life on the planet must have begun (or at least become recognizable as such) with some kind of free-living single cell. A transition must then have been made from isolated single cells to a multicellular organization, probably through the intermediary of a colonial stage (examples of such colonial organisms which still exist today are a staple of freshman biology courses). These multicellular organisms must have become highly diversified, and ultimately some of them (perhaps a group related to the modern annelid worms or to the coelenterates) started on the path which led to the backbone. Palaeontological evidence suggests unavoidably that the earliest vertebrates were marine, and hence that the earliest vertebrate members of the phylogenetic sequence leading to the mammals

and man possessed gills. The earliest vertebrates also lacked a true skeleton made of bone, but they possessed one made of cartilage.

Now look at the ontogenetic sequence of a typical vertebrate. Ontogeny begins with a single cell (the fertilized egg), just as the phylogenetic sequence did. Repeated divisions of that fertilized egg lead to a colonial type of organization of similar cells (blastula). Further differentiation leads to a form reminiscent of an invertebrate organism. At a somewhat later stage, the developing organism, or embryo, possesses a preliminary skeleton built of cartilage. It also possesses a set of branchial or visceral arches, similar to the gills of a typical marine vertebrate (indeed, these develop into actual gills in the sharks and bony fish).

It is perfectly obvious that the parallels between ontogeny and phylogeny, revealed by the briefest of surveys of the two processes, and which can be elaborated in great detail, indicate that some rather deep biological principle underlies them both. The nature of that principle was formerly a much-debated question, and it will be helpful to briefly review some of the pertinent arguments.

Probably the best-known view of the ontogenetic-phylogenetic parallelism, first elaborated by Haeckel[3] and his followers, held that the ontogenetic sequence is a recapitulation of the corresponding phylogenetic sequence. This view, if correct, has many profound implications which Haeckel and others were not slow to point out. In the first place, the recapitulation hypothesis, or the biogenetic law, as it came to be known, provides a direct proof of evolution (i.e. of a phylogenetic sequence underlying every extant organism). Every organism certainly possesses an ontogeny; if ontogeny must recapitulate phylogeny, then phylogeny must exist for each organism also.

More important, however, was the implication that an empirical study of the ontogenetic sequence would make possible a direct study of phylogenesis. In other words, if the biogenetic law were valid, it would follow that phylogenesis could be studied by the same direct quantitative means by which one studied ontogenesis. Direct insight into phylogenesis, which might otherwise be impossible to obtain because of the fragmentary nature of the fossil record, thus becomes immediately possible, *if* one learns how to read the record written by recapitulation.



These hopes, however, have not proved to be well-founded, for reasons which will be important to later discussion. In its earliest form, the biogenetic law required that, if an advanced form *B* evolved from a simple form *A*, then the ontogenetic stages corresponding to the intermediary forms between *A* and *B* are simply added in sequence onto the ontogenetic sequence of *A*. But this is rarely, if ever, observed. Moreover, it became clear that ontogenesis often shuffled, inverted, or simply omitted stages corresponding to known phylogenetic steps. Conversely, ontogenetic sequences were found to contain stages divorced from any phylogenetic interpretation whatsoever. A whole new terminology was created, and the biogenetic law correspondingly modified, to try to render these situations consistent with the basic recapitulation hypothesis, but the situation grew gradually more and more unwieldy, and was finally given up, at least as a theoretical principle[4].

At the risk of oversimplifying these complex matters, one may note two characteristics of ontogenesis which made things difficult for the biogenetic law:

(1) The rather free omissions and intercalations into ontogenesis of stages with dubious phylogenetic significance, and the pronounced shuffling of such stages which could take place without greatly affecting the end-result of the ontogenetic process;

(2) More generally, the flexibility of what one might call 'ontogenetic time' as compared with the fixed, precise sequence characteristic of 'phylogenetic time.'

The difficulties encountered by the biogenetic law led other biologists to seek for another explanation of the parallels, which no one denied, between ontogeny and phylogeny. A competing view was put forward by Lillie, and others[5], which attempted to avoid those difficulties. According to these views, the fundamental defect of Haeckelian recapitulation was that phylogenesis was regarded as a sequence of adult forms, and ontogenesis as an 'accelerated repetition' of these adult forms. Rather, it was held, phylogenesis must be considered as a sequence of complete ontogenies. In this view, then, the modification of some stage in the ontogeny of an initial organism *A* leads to alterations in the adult form of *A*, *and* in the ontogenesis of all descendants of *A*. Thus it is not too much amiss to regard this view as an inverse of

that of Haeckel, and to paraphrase it as 'phylogeny recapitulates ontogeny'. As Garstang[6] puts it:

> '[Ontogeny] reproduces successive grades [of ancestral differentiation], not because successive adult types have been included in it, but because each ontogeny is a modification, within limits, of its predecessor; and by those predecessors the phyletic chain of adults was organized and equipped.'

This view suffers from its own difficulties. For instance, Haeckel could interpret the embryonic transition from the cartilaginous to the bony skeleton to be a recapitulation of the phyletic transition from the sharks to true bony fish. However, the followers of Lillie's view had to claim that cartilage is an obligatory precursor of bone, and that developing organisms, either in ontogeny or phylogeny, can have no way of manufacturing bone except through the intermediary of cartilage. Likewise they must assert that there is no way to complete the ontogenesis of terrestrial vertebrates except through an intermediary stage involving the famous branchial arches and gill slits[7]. The myriad of such claims which have to be made do not lend great weight to this view. It should also be noticed that Lillie's view, being independent of specific mechanisms, utterly lacks the predictive power which was one of the main attractions of the biogenetic law.

It will be seen, however, that a view very similar to that of Lillie emerges as a consequence of the general considerations which are in the process of being developed.

## 4.4 Ontogeny and Natural Selection

There is one more important point to make before it can be seen clearly where the discussion of the past few sections is leading. This point involves perhaps the major distinction between the ontogenetic and the phylogenetic sequences, and is simply this: that whereas phylogenesis is a process governed entirely by natural selection, ontogenesis is designed to be largely or completely free of direct selection pressure. This simple point, properly understood, is enough to explain many of the similarities, and also the differences, between the two types of development. It also allows one to bring the mathematical tools to bear on



important problems which otherwise could not be approached directly.

Let us first clarify the meaning of the assertion that selection pressure is absent during much, or all, of ontogenesis. The point is that selection pressure is always defined in terms of a differential reproductive capacity among competing forms, as shown in Chapter 1. If there are no competing forms sharing the same niche, then by definition there is no pressure of selection in that niche. It then remains to argue that, during much, or all, of ontogenesis, the developing organism does not share a niche with any other organism.

Among higher organisms, the insulation of a developing embryo from direct competition is particularly striking. In the mammal, several embryos generally develop in the same uterus simultaneously, but each one typically has its own set of embryonic membranes, its own placenta and its own umbilical attachment. Among birds and egg-laying reptiles, each embryo is actually physically insulated from any contact with the external environment, and is provided with its own private food reservoir. The same idea is manifested throughout the remainder of the plant and animal kingdoms, although in simpler forms the mechanisms to ensure this insulation are less elaborate and perhaps less effective.

The assertion about the absence of selection from ontogenesis has been qualified because of the many species which possess free-living larval forms; such forms do of course compete directly. But most of the truly significant ontogenetic changes occur before such a form is attained, and what has been said applies to all those ontogenetic stages which precede the larval phase.

*Exercise*

Discuss the selective value of the suppression of competition during ontogenesis. (Hint: what effect does minimizing competition between developing young have on fecundity?)

If, then, one is prepared to accept that direct selection pressure plays a minimal, and often negligible, role during ontogenesis, then the points of the argument may be brought together.

### 4.5 Ontogeny and Optimality

In Chapter 1 and the preceding sections, the following points were made:

(1) That selection pressure is the major determinant of structure *among competing organisms*.

(2) That the effect of selection pressure is described mathematically as an optimization of biological structure.

(3) That phyletic sequences involve competing forms, and hence by (2) are governed by principles of optimality.

(4) That selection pressure plays at best a small part during ontogenesis, and hence there is no reason to suppose that ontogeny is directly governed by optimality principles of any kind.

(5) That many close similarities nevertheless exist between the ontogenetic and phylogenetic sequences.

Indeed, with respect to (4) above, many instances of what appear to be superfluous structural stages are known, and have been cited in the literature[8], precisely as an antidote to that view of nature as a machine for grinding out optimal structures.

The problem is whether mathematical tools, based on principles of optimality, may or may not be applied to problems of ontogenetic development. The observation (5) indicates that they can be applied, provided that the relation between ontogeny and phylogeny can be elucidated. For since, by (3), phylogenetic transitions can be so treated, that relation will allow one to transfer, at least in a formal sense, the conclusions of phylogenetic studies to the realm of ontogeny.

On the other hand, (4) seems to indicate precisely the opposite: that there is no reason to suppose that any of the structures arising during ontogenesis are in any sense optimal. Thus the application of optimality techniques to ontogenesis faces two major problems: (*a*) to determine how it is possible that (4) and (5) can be simultaneously true, and (*b*) to clarify the relation between ontogeny and phylogeny.

Both these problems are, to a large degree, clarified by recalling once again the absence of selection pressure during ontogenesis. The absence of selection means that, within wide limits, any kind of adventitious or *ad hoc* alteration of the ontogenetic sequence will survive ontogenesis, because none of these alterations is selected against during ontogenesis itself. Once past ontogenesis,



some of these alterations will give rise to adult forms possessing a selective advantage; these forms will be conserved. Some will be absolutely neutral, i.e. cause no alteration in adult form. Others will convey a selective disadvantage, and will tend to be eliminated. In other words, any modification of the ontogenetic sequence which do not decrease the fecundity of the adult form will tend to be conserved indefinitely. This means precisely that phyletic structures of adult organisms, which are at any time optimal, tend to carry the ontogenetic stages which gave rise to them; and further, these stages may persist in the ontogenesis of all subsequent forms, because there is no direct way to select against them during ontogenesis itself.

This view is in some ways very close to that of Lillie, described above. However, it avoids some of the difficulties of that view. Most important, it enables the machinery of optimality techniques to be applied to many kinds of ontogenetic problems, despite the fact that selection does not play a direct role in ontogenesis. Full advantage of this licence is taken in Chapters 5 and 6.

## Notes to Chapter 4

1. As with all biological generalizations, this assertion admits of counter examples which are of some interest. The larvae of certain insects and amphibia can acquire reproductive capacity and propagate indefinitely. Two related processes seem to be involved; one, called *paedogenesis*, is the acquisition of reproductive capacity by a larval form; the other, called *neoteny*, is the retention of larval characteristics in the adult state. The best-known examples of paedogenesis are the axolotl, which is a larval salamander, and the glow-worm. Modern examples of paedogenesis and neoteny seem to be without phylogenetic significance, but it has been suggested that these processes might have played an important role in the evolution of true vertebrates from more primitive chordate ancestors (which normally possess a notochord only in larval stages).

2. An interesting example of selection, which may be cited in this context, is *industrial melanism*. See, for example, H. B. D. Kettlewell, 'Investigations on the Evolution of Melanism in Lepidoptera', *Proc. roy. Soc.*, B, CXLV (1956) 297.

3. E. Haeckel, *Generelle Morphologie der Organismen*, Berlin, 1866. Perhaps the most accessible work by Haeckel in English is *The Evolution of Man*, transl. J. McCabe, Peter Eckler, New York, 1906. Of course, ideas resembling various facets of the Haeckelian view had been elaborated long before; for an excellent historical review see A. W.